

PREDICCIÓN DE LA DEMANDA TURÍSTICA EN LA REGIÓN NORTE DE PORTUGAL

Paula Odete Fernandes* - e-mail: pof@ipb.pt
Luis Borges González** - e-mail: borge@eco.uva.es
Francisco José Terroso Cepeda* - e-mail: cepeda@ipb.pt

* - ESTiG - Instituto Politécnico de Bragança, Portugal.

** - Facultad de Ciencias Económicas y Empresariales, Universidad de Valladolid, España.

RESUMEN

Al igual que otros países, en Portugal el turismo es una actividad que genera riqueza y valor añadido y cada día va asumiendo un papel cada vez más relevante en el seno de su economía. Por tanto, la predicción de la demanda turística debe de formar parte integrante de la gestión diaria de cualquier empresa/organización, pública o privada, particularmente ligadas a este sector, cuestionando y debatiendo las hipótesis subyacentes a las predicciones realizadas y utilizando los conocimientos articulados con todo el análisis del proceso de gestión.

En este contexto, resulta importante efectuar un estudio de predicción encaminado no sólo a determinar la evolución de la demanda turística para la Región Norte de Portugal, sino también a comparar los modelos desarrollados y que pueden ser susceptibles de ser utilizados como instrumentos de predicción de la demanda turística.

La metodología con lo que se llevo a cabo este análisis fue la de *Box-Jenkins*, aplicada sobre la serie "*Pernoctaciones Mensuales en la Región Norte de Portugal*", registradas en el período de Enero de 1987 a Diciembre de 2002, puesto que es una de las variables que mejor explica la demanda efectiva.

Se constató que modelo $ARIMA(0,1,1) \times (0,1,1)_{12}$, obtenido con la metodología aplicada, fue el que presento calidades estadísticas satisfactorias, que mejor se ajusto al comportamiento de la serie de referencia y que permitió realizar previsiones de la demanda turística para la Región Norte de Portugal, en cada una de las metodologías.

Palabras Clave: Estacionariedad, Estacionalidad, Modelos ARIMA y Predicción.

1. INTRODUCCIÓN

El estudio de la demanda turística es fundamental para la organización de los sectores y entidades relacionadas con el turismo para que puedan desarrollar medidas estratégicas que sean capaces de obtener el máximo rendimiento de las situaciones emergentes. Por ello, dedicamos este artículo a estudiar y predecir el comportamiento de la demanda turística para la Región Norte de Portugal.

La metodología con lo que vamos a llevar a cabo este estudio es la de *Box-Jenkins* aplicada sobre la serie "*Pernoctaciones Mensuales en la Región Norte de Portugal*", puesto que es una de las variables que mejor explica la demanda efectiva.

Para la metodología se realizará un análisis temporal de referencia, recurriendo al conocimiento de sus valores en el pasado. Este análisis tiene como objetivo principal obtener el modelo que mejor se ajuste al comportamiento de la serie en análisis y utilizarlo para hacer predicciones. Con el fin de evaluar las predicciones del modelo, se ha reservado una parte de los datos para poder medir la calidad de cada predicción.

2. MODELO UNIVARIANTE DE *BOX-JENKINS*

Box y *Jenkins* han desarrollado modelos estadísticos que tienen en cuenta la dependencia existente entre los datos. Cada observación en un instante de tiempo se modeliza como una función lineal de los valores anteriores.

El análisis consiste en la obtención de un modelo explícito. Estos modelos se conocen con el nombre genérico de *ARIMA* (*Autoregressive Integrated Moving Average*), autorregresivo integrado de media móvil, el cual deriva de sus tres componentes: autorregresivo (AR), integrado (I) y de medias móviles (MA). El modelo *ARIMA* permite describir un valor como una función lineal de datos anteriores y errores debidos a la casualidad. Puede incluir, además, un componente cíclico o estacional. El objetivo que se persigue es la obtención, a partir de unos datos, un modelo adecuado al comportamiento de los datos y que sea lo más parsimonioso posible. Es decir, debe de contener todos los elementos necesarios para describir los datos, pero los mínimos posibles para describir el fenómeno. *Box* y *Jenkins*, recomiendan un mínimo de 50 observaciones en la serie temporal (Murteira *et al.*, 1993; Makridakis & Hibon, 1997). Siguiendo esta metodología, es posible encontrar diferentes modelos algebraicos que explican el comportamiento de los datos y que son adecuados para realizar predicciones a corto plazo, incluyendo el caso de series que contengan variaciones estacionales (Witt & Witt, 1995; Greene, 2000).

El proceso de estudio y de análisis de una serie temporal comprende por tanto una serie de fases que comienza con el estudio de la estacionalidad de los datos, o su transformación si fuera necesario. A continuación, teniendo en cuenta el teorema de descomposición de Wold¹ (1938), podemos sintetizar que la metodología de *Box-Jenkins* consiste en la búsqueda de dos polinomios de retardos, uno que actúa sobre las observaciones y que denominamos parte

¹Este resultado establece que cualquier proceso estocástico estacionario puede ser representado de manera única como la suma de dos procesos mutuamente incorrelacionados, siendo el primero determinista y el segundo un proceso lineal ($MA(\infty)$).

$$Y_t = D_t + X_t, \text{ el proceso lineal } X_t, \text{ se define a partir de un ruido blanco como:}$$

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i e_{t-i} \quad \text{donde} \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty.$$

autorregresiva del modelo, y otro que actúa sobre un proceso de ruido blanco y que se denomina componente de medias móviles. Debido al esquema seguido en el planteamiento del modelo y su resolución operativa, estos métodos permiten la incorporación de análisis de carácter estacional, el aislamiento de la componente tendencia e, incluso, profundizar las interrelaciones entre estas componentes que se integran en la evolución de las series en estudio (Parra & Domingo, 1987; Chu, 1998).

La modelización de una serie temporal, por tanto, consiste en derivar un Modelo ARIMA que se ajuste al conjunto de datos dados. Las fases para llevar a cabo esta tarea, propuestas por G. Box y G. Jenkins en 1970, son las siguientes (Murteira *et al.*, 1993; Kohzadi *et al.*, 1996; Makridakis & Hibon, 1997; Johnston & DiNardo; 2000):

A continuación se va describir series temporales estacionarias bajo el enfoque de *Box-Jenkins*. El proceso básico sobre el que se basa toda la teoría es el de ruido blanco y a partir de él se construyen los diferentes modelos ARMA.

Un proceso Y_t se considera generado por un modelo ARMA(p,q), si puede expresarse como:

$$Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad [1]$$

siendo e_t un ruido blanco, es decir una variable aleatoria de media cero y varianza constante.

El resto de los elementos son los parámetros del modelo. La expresión anterior se puede escribir en función de un polinomio de retardos (siendo B el operador de retardos), en la forma:

$$\phi_p(B)Y_t = \theta_q(B)e_t \quad [2]$$

donde,

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad [3]$$

y

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad [4]$$

La estacionariedad exige que cumpla un doble requisito en la práctica: una media aproximadamente constante en el tiempo y una varianza o dispersión también constante. Es decir, que si se divide una serie en subperíodos arbitrarios, la media y la varianza sean aproximadamente iguales en cada uno de ellos. En forma gráfica, una media constante supone la no existencia de tendencia (una línea paralela al eje de las abcisas). Una varianza constante corresponde a un gráfico donde las oscilaciones alrededor de la media son semejantes, lo que técnicamente es conocido como homocedasticidad (Pulido, 1989).

En un modelo ARMA las condiciones de estacionariedad dependen sólo de la parte autorregresiva del modelo. Las raíces del polinomio de retardos de orden p , $\{\phi_p(B)\}$, deben estar fuera del círculo de radio unitario. Por otro lado, las condiciones de invertibilidad (el modelo se puede representar como un MA(∞)) afectan sólo a la parte de media móvil del modelo, y requieren la misma condición para las raíces del polinomio de retardos de orden q , $\{\theta_q(B)\}$. Las condiciones de estacionariedad y de invertibilidad son independientes. Por tanto, el proceso ARMA(p,q) es estacionario si lo es su componente autorregresiva, y es invertible si lo es su componente de medias móviles.

Los modelos ARMA(p,q), según se desprende de su definición, son modelos estacionarios. Sin embargo, muchas series económicas, como se desprende de su representación gráfica, presentan tendencia, creciente la mayor parte de las veces, o decreciente. Una de las formas más elementales para eliminar la tendencia consiste en aplicar diferencias sucesivas sobre la serie original. Si después de tomar una diferencias, la serie aún no es estacionaria en media se sigue diferenciando hasta obtener una serie que sea estacionaria en media. La serie se transformará en la forma:

$$\nabla^d Y_t = (1 - B)^d Y_t = W_t \quad [5]$$

de forma que el proceso W_t es estacionario.

Ya se ha comentado una forma sencilla de eliminar la tendencia mediante la diferenciación sucesiva de la serie original. Sin embargo, esta transformación no elimina el fenómeno de la heterocedasticidad, que la varianza no permanezca constante. Esta característica, muy estudiada en econometría y que presentan muchas series, es imprescindible corregirla para que la serie sea estacionaria. La solución viene dada por una transformación de los datos, la transformación de *Box-Cox*, que es (Murteira *et al.*, 1993):

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}, & \text{para } \lambda \neq 0 \\ \ln Y_t, & \text{para } \lambda = 0 \end{cases} \quad [6]$$

En la práctica, es la transformación logarítmica la que, en la gran mayoría de las circunstancias, transforma las series en homocedásticas.

El modelo ARIMA(p,d,q), donde: p corresponde al orden del proceso autorregresivo (AR), d es el número de diferencias o integraciones y q corresponde al orden del proceso de medias móviles (MA), puede escribirse como:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d Y_t = (1 - \theta_1 B - \dots - \theta_q B^q) e_t \quad [7]$$

o de forma abreviada:

$$\phi_p(B) \nabla^d Y_t = \theta_q(B) e_t \quad [8]$$

Otra posible fuente de no estacionariedad en las series económicas reales la constituye la estacionalidad, fundamentalmente cuando los datos se corresponden con una frecuencia mayor que la anual (datos mensuales o trimestrales). Las series de este tipo, presentan picos que aumentan o disminuyen los valores de la serie siempre en los mismos periodos. Este comportamiento se elimina tomando diferencias estacionales. La serie transformada será igual a:

$$\nabla_s^d = (1 - B^s)^d Y_t = W_t \quad [9]$$

donde W_t sigue un modelo ARMA(p,q).

Las aplicaciones de los modelos ARIMA se suelen realizar con datos trimestrales, mensuales o incluso semanales, diarios u horarios. Es decir, en un contexto a corto plazo. A tales efectos, los modelos ARIMA tratan de captar el comportamiento estacional de una manera idéntica al tratamiento de la componente regular (o no estacional) de la serie.

En las aplicaciones no es frecuente trabajar con un solo modelo ARIMA (p,d,q) , sino con un producto de los modelos: $ARIMA(p,d,q)(P,D,Q)_s$ donde el primero corresponde a la parte regular y el segundo a la estacional.

En general, un modelo completo $ARIMA(p,d,q)(P,D,Q)_s$ corresponde a cualquiera de las siguientes expresiones:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)e_t \quad [10]$$

o

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla^D Y_t = \theta_q(B)\Theta_Q(B^S)e_t \quad [11]$$

Aún, el modelo [10] es susceptible de ampliación con m polinomios autorregresivos estacionales y n polinomios de medias móviles estacionales, capaz de describir una situación en que existan componentes estacionales o cíclicas con distintos períodos, adoptando la siguiente expresión:

$$\begin{aligned} \phi_{p(B)} \left\{ \prod_{i=1}^m \Phi_{P_i}(B^{S_i}) \right\} (1-B)^d \left\{ \prod_{i=1}^{m^*} (1-B^{S_i})^{D_i} \right\} Y_t = \\ = \theta_0 + \theta_q(B) \left\{ \prod_{i=1}^n \Theta_{Q_i}(B^{S_i}) \right\} e_t. \end{aligned} \quad [12]$$

Las fases para derivar un Modelo ARIMA que se ajuste al conjunto de datos dados, propuestas por G. Box y G. Jenkins en 1970, son las siguientes (Murteira *et al.*, 1993; Kohzadi *et al.*, 1996; Makridakis & Hibon, 1997; Johnston & DiNardo; 2000):

1ª Identificación del modelo, a través de una muestra suficientemente grande de datos relativamente recientes. Esto es, determinar el tipo de modelo más adecuado para la serie objeto de estudio;

2ª Estimación: estimación de sus parámetros, una vez bien especificado el modelo.

3ª Validación: evaluación del ajuste antes referido y, si es necesario, intentar encontrar un nuevo modelo que describa mejor el fenómeno en causa.

Los modelos que superan la etapa anterior pueden ser utilizados para uno de los fines básicos de la metodología de *Box-Jenkins* que es la obtención de predicciones.

Si en cada una de las últimas dos etapas el modelo estimado no resulta satisfactorio, es necesario volver a la fase de identificación y recorrer nuevamente las fases del método, que esquemáticamente viene relegado en la Figura 1.

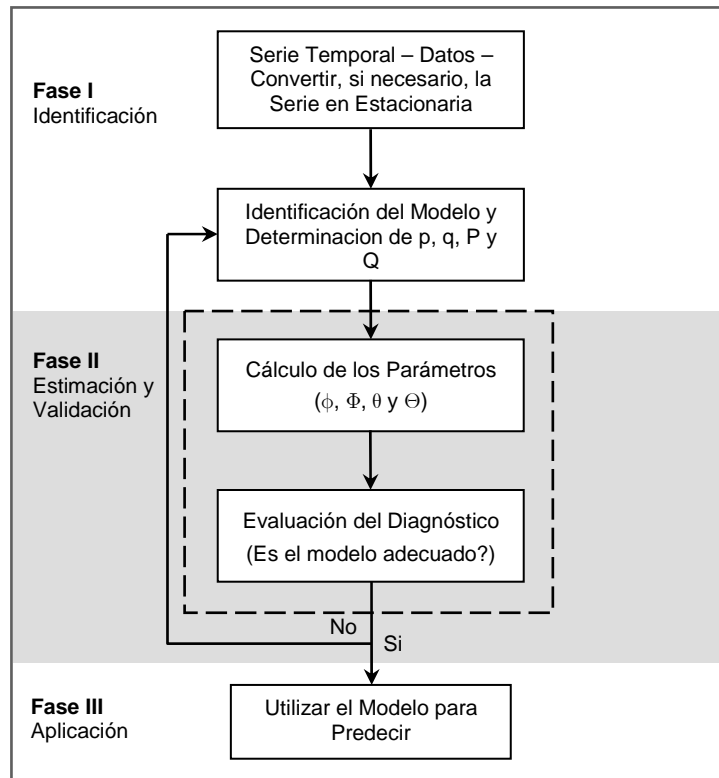


Figura 1: Esquema General de la Metodología de *Box-Jenkins*.

Fuente: Adaptado de MAKRIDAKIS y HIBON; (1997:149).

3. DEMANDA TURÍSTICA EN LA REGIÓN NORTE DE PORTUGAL

3.1. Presentación y Estacionariedad de la Serie

A lo largo del análisis que a continuación se presenta, la serie “*Pernoctaciones Mensuales de la Región Norte de Portugal*” pasará a designarse por SO_t (Serie Original).

En la primera fase de identificación, el primer paso es comprobar la estacionariedad de los datos, o en su defecto, realizar transformaciones que conviertan la serie en estacionaria. El primer paso, es una representación gráfica de la serie SO_t , obtenida en el Instituto Nacional de Estadística (INE) Portugués. Los datos están recogidos en el Anexo A - Tabla A.1, y la Figura 2 se corresponde con su representación gráfica.

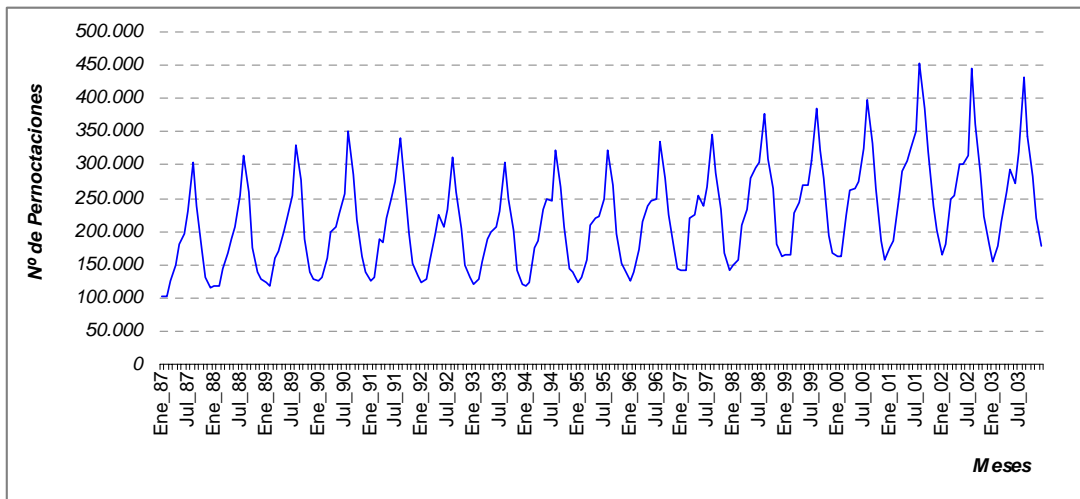


Figura 2: Cronograma de la serie SO_t , para el período de 1987:01 a 2003:12.

La observación de la Figura 2 denota un ligero crecimiento en la tendencia de la serie, un posible crecimiento en el tamaño de las oscilaciones y la clara existencia de un componente estacional ('picos' regularmente espaciados en el tiempo, máximos en los meses de Verano y mínimos en los meses de Invierno). Es decir, los valores de las pernoctaciones vienen en función de la época del año. Para asegurar la estabilidad de la varianza en la serie, se construyó el gráfico de Media-Rango, con las medias frente a las desviaciones típicas de los datos de cada año. Su representación en la Figura 3 confirma el crecimiento de la varianza.

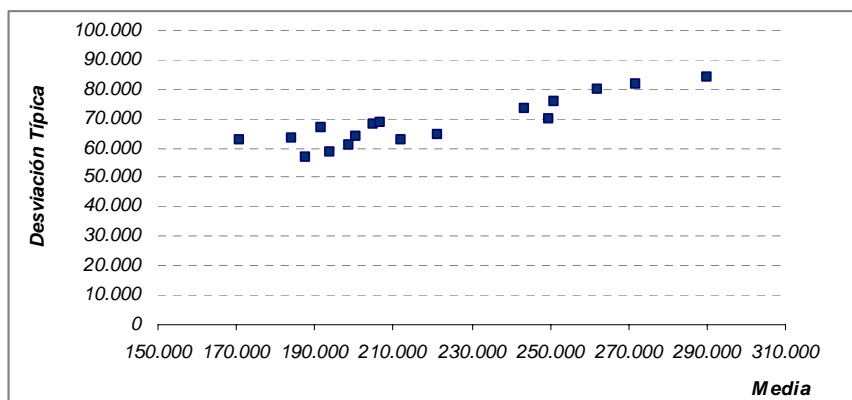


Figura 3: Grafico Media-Rango.

Con el fin de estabilizar la variancia de la serie SO_t , se ha transformado la serie original aplicando el logaritmo natural, $LSO_t = \ln(SO_t)$. La serie resultante sigue siendo de manifiestamente no estacionaria pero más estable en variancia, notándose aún una tendencia creciente y la existencia de movimientos periódicos (ver Figura 4). Por ello, en la continuación del estudio, el análisis tendrá como base la serie original transformada $[LSO_t]$ y el período de 1987:01 a 2002:12.

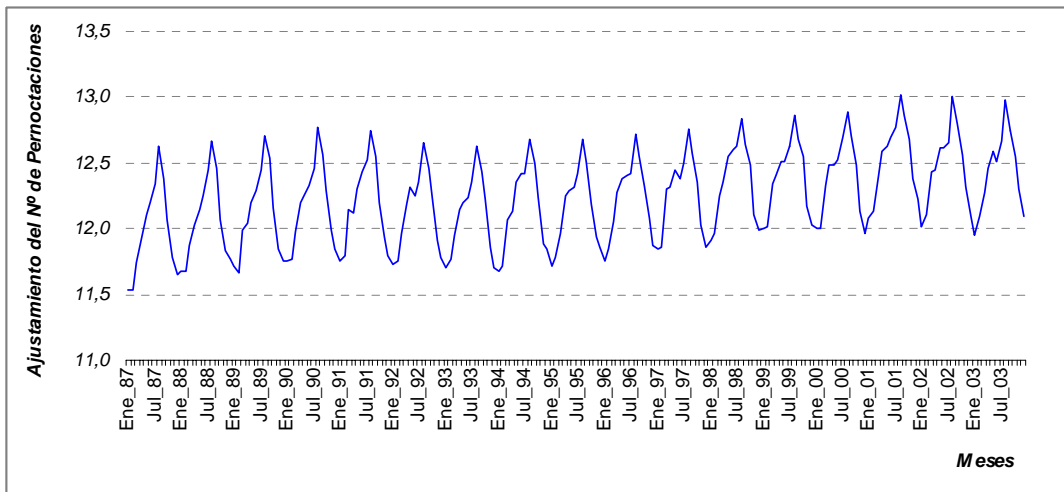


Figura 4: Cronograma de la serie LSO_t , para el período de 1987:01 a 2003:12.

Después del análisis de las FAC^2 y FAP^3 estimadas de la serie, se llega a conclusiones semejantes a la anterior. La FAC (Figura 5), toma la forma de una senoide amortecida, no decreciendo hacia cero a medida que K aumenta, lo que confirma que se está ante un proceso no estacionario. La estacionalidad también es evidente en la observación de la FAC , que presenta autocorrelaciones positivas elevadas para valores de K múltiples de 12. Los picos ocurren todos los años en los meses de Agosto, sugiriendo una estacionalidad de período 12, puesto que se trabaja con valores mensuales.

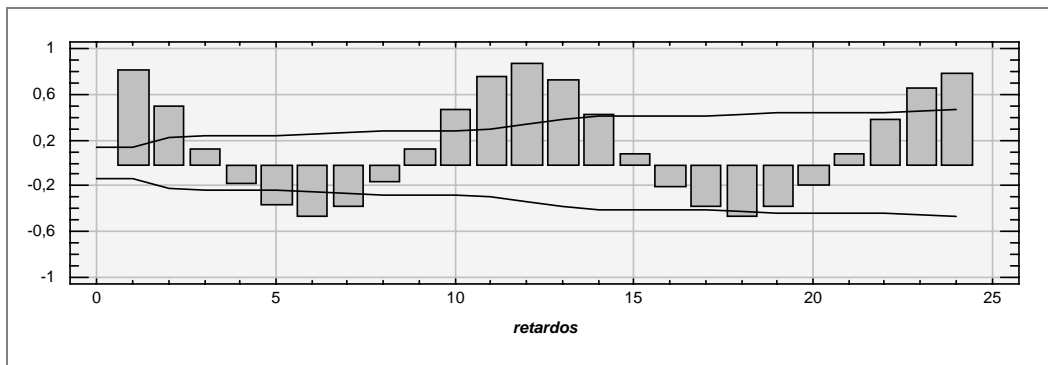


Figura 5: FAC estimada de la serie LSO_t .

El paso siguiente, para remover la tendencia y la componente estacional, es tomar diferencias. Tales transformaciones, a pesar de que conserven la estructura general de la serie, ejercen efectos considerables sobre el conjunto de datos que pueden permitir su estudio más cómodo, alterando su escala (eventualmente disminuyendo su amplitud), atenuando asimetrías, eliminando posibles *outliers*, disminuyendo residuos y alcanzando, finalmente, los objetivos en causa: estabilizando varianzas y linealizando tendencias (Otero; 1993).

² Función de Autocorrelación.

³ Función de Autocorrelación Parcial.

Se consiguió convertir la serie en estacionaria, tanto en media como en varianza, después de la aplicación de una diferenciación simple, $[\nabla Y_t = Y_t - Y_{t-1} = (1-B)Y_t]$ y de una diferencia estacional $[\nabla_s Y_t = Y_t - Y_{t-s} = (1-B^s)Y_t]$. Es decir, se tomaron diferencias sucesivas y diferencias entre las observaciones separadas por el periodo estacional (12, dado que la información es mensual), transformándose la serie original en una nueva serie. El resultado de la nueva serie, que servirá de base para estudiar las fases sucesivas, es: $(1-B^{12})(1-B)LSO_t$, cuyo cronograma viene reflejado en la Figura 6.

Es conveniente minimizar las diferenciaciones de los datos para evitar el efecto de la sobre-diferenciación⁴, pues diferenciar origina aumento de la varianza del error de la predicción (Murteira *et al.*, 1993; González, 1999).

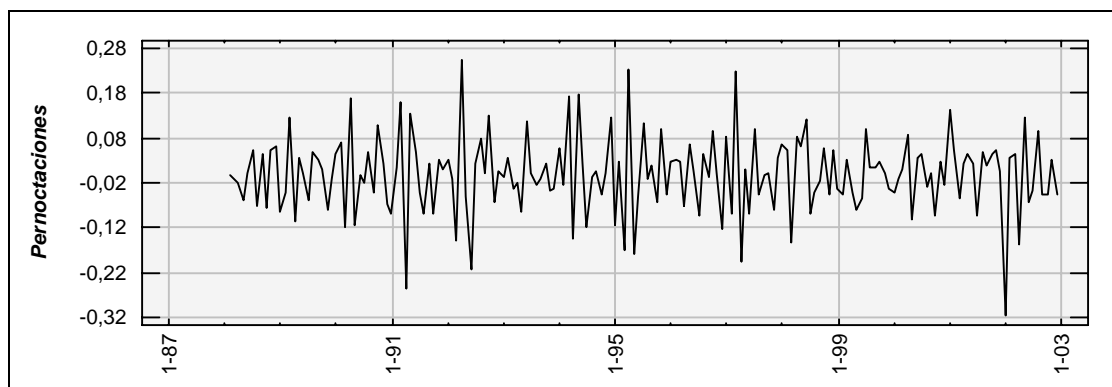


Figura 6: Cronograma de la serie $(1-B^{12})(1-B)LSO_t$.

3.2. Identificación del Modelo

El proceso de identificación está basado tanto en la FAC como en la FAP. Cada modelo lineal estacionario construido teóricamente, tiene asociado una función de autocorrelación y una función de autocorrelación parcial asociada a diferentes retardos. Los gráficos correspondientes a estas funciones teóricas de distintos modelos, en general, se corresponden solamente con la parte regular de la serie, aunque son fácilmente generalizables a la parte estacional (Otero, 1993; Pulido, 1989), y se pueden encontrar en numerosas fuentes, entre ellas las citadas anteriormente.

En la Tabla 1, se presenta el comportamiento de los coeficientes en las funciones FAC y FAP en los procesos AR(p), MA(q) y ARMA(p,q) que van ayudar a identificar el modelo, sabiendo también que el comportamiento de las FAC y FAP en los procesos estacionarios tienden hacia cero a medida que aumentan los retardos.

⁴ También designado por *overdifferencing*.

Tabla 1: Cuadro resumen del comportamiento de las FAC y FAP en procesos AR, MA y ARMA.

PROCESO	FAC	FAP
AR(p)	Exponenciales atenuados u ondas sinusoidales, todos los coeficientes son no nulos que decrecen con el retardo.	Los p primeros coeficientes significativos solamente y los restantes cero.
MA(q)	Los q primeros coeficientes significativos solamente y los restantes cero.	Exponenciales atenuados u ondas sinusoidales, todos los coeficientes son no nulos que decrecen con el retardo.
ARMA(p,q)	Decrecimiento para cero de forma exponencial u ondas sinusoidales amortiguadas.	Decrecimiento para cero de forma exponencial u ondas sinusoidales amortiguadas.

Fuente: Adaptado de MURTEIRA *et al.*; (1993:100).

La identificación de las componentes estacional y no estacional, se realiza separadamente, recurriendo a la información presentada en la Tabla 1. Observando las funciones de autocorrelación y autocorrelación parcial estimadas de la serie del estudio, Figuras 7 y 8, respectivamente, hay que decidir:

- a) los orden de la componente regular (p,q), basándose en la evolución de los primeros coeficientes ($lags, K=1,2,3,\dots$);
- b) el orden de la componente estacional (P,Q) observando el comportamiento de los coeficientes para los retardos estacionales: 12 y 24 ($lags, K=S,2S,3S,\dots$).

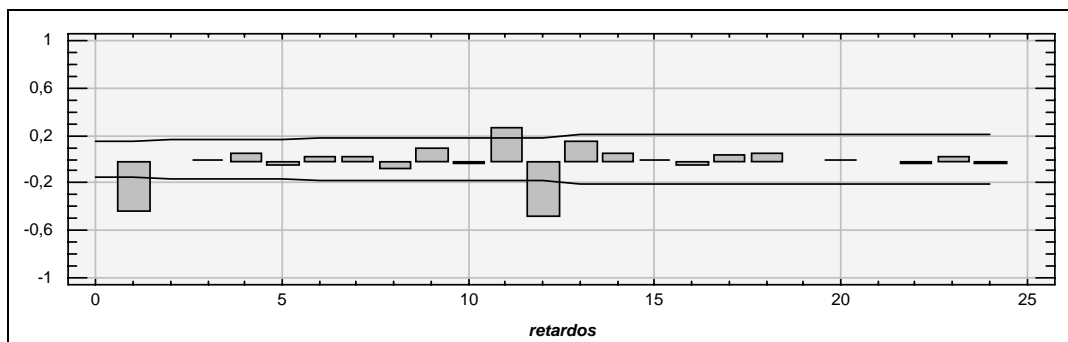


Figura 7: FAC estimada de la serie después de la diferenciación simple y estacional.

El análisis de las FAC y FAP (Figuras 7 y 8, respectivamente) estimadas de la serie estacionaria, considerando un intervalo de confianza al 95%, parecen sugerir un proceso, para la componente no estacional, ARMA (0,1) ya que la primera estimación de la FAC es significativa, decayendo bruscamente para cero a partir del primer retardo y los valores iniciales de la FAP son significativos y decayendo de una forma exponencial.

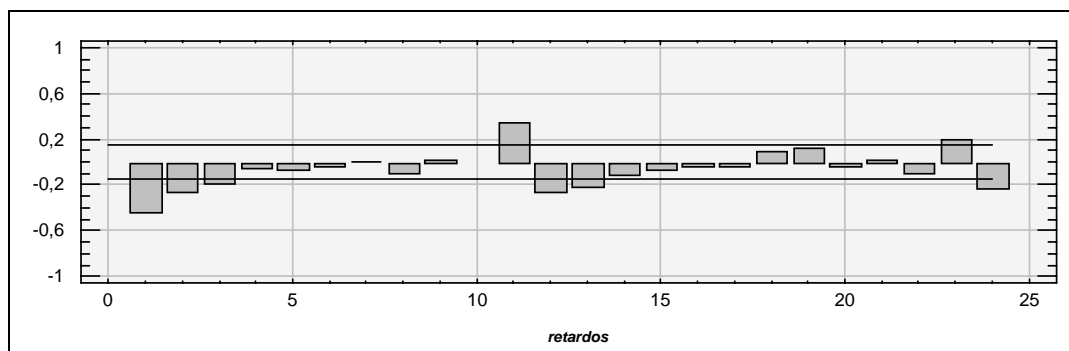


Figura 8: FAP estimada de la serie después de la diferenciación simple y estacional.

Con relación a la componente estacional, las FAC y FAP estimadas, sugieren igualmente un proceso ARMA (0,1) estacional (SARMA), atendiendo a los valores de la FAC estimada en el retardo 12 y 24 (el primero es significativo, mientras que el segundo no lo es), y los valores de la FAP para los mismos retardos, ambos significativos. Por tanto, el modelo sugerido es del tipo $ARIMA(0,1,1)\times(0,1,1)_{12}$, que será designado por Modelo M_1 . Sin embargo, existen otros modelos similares que pueden entrar en competencia con el anterior, por lo que se va a tener en cuenta también en el estudio los modelos $ARIMA(1,1,1)\times(1,1,1)_{12}$ y $ARIMA(1,1,1)\times(0,1,1)_{12}$, que designaremos Modelos M_2 y M_3 , respectivamente. Estos modelos, deben ser considerados como los mejores candidatos para ajustar el comportamiento de la serie. Para elegir entre ellos, se aplican los criterios de selección que veremos a continuación.

3.3. Estimación, Validación y Selección del Modelo

Una vez identificados los modelos ARIMA, que mejor se han ajustado a la serie, hay que determinar los valores de los parámetros de las funciones lineales que los definen. El método utilizado para la estimación de los parámetros ϕ e θ es el de mínimos cuadráticos, obteniéndose los siguientes resultados:

Tabla 2: Estimaciones obtenidas para los coeficientes de los modelos.

MODELOS	PARÁMETROS	RETARDOS	COEFICIENTE	DESVIACIÓN TÍPICA	RATIO t	P-VALUE	MEDIA DE LOS RESIDUOS	VARIANCIA DE LOS RESIDUOS
M_1	Medias Móviles	1	0,628237	0,0594336	10,5704	0,000000	-0,00187	0,0031759
	Medias Móviles	12	0,769101	0,0423414	18,1643	0,000000		
M_2	Autorregresivo	1	0,13061	0,112691	1,15901	0,248032	-0,00158	0,0030962
	Medias Móviles	1	0,725991	0,0772333	9,39998	0,000000		
	Autorregresivo	12	-0,260857	0,100318	-2,6003	0,010110		
	Medias Móviles	12	0,64362	0,0708315	9,08663	0,000000		
M_3	Autorregresivo	1	0,195578	0,11049	1,77011	0,078441	-0,00152	0,0031498
	Medias Móviles	1	0,758743	0,0735282	10,3191	0,000000		
	Medias Móviles	12	0,796259	0,0410751	19,3855	0,000000		

Cada uno de los modelos estimados vienen dados por las siguientes expresiones:

$$M_1 = \nabla \nabla_{12} LSO_t = (1 - 0,628237B)(1 - 0,769101B^{12})e_t \quad [13]$$

$$\begin{aligned} M_2 &= (1 - 0,13061B)(1 + 0,260857B^{12})\nabla \nabla_{12} LSO_t = \\ &= (1 - 0,725991B)(1 - 0,64362B^{12})e_t \end{aligned} \quad [14]$$

$$M_3 = (1 - 0,195578B)\nabla \nabla_{12} LSO_t = (1 - 0,758743B)(1 - 0,796259B^{12})e_t \quad [15]$$

El análisis del significado estadístico de los coeficientes estimados para el modelo M_1 , revela que los mismos son significativamente diferentes de cero, a un nivel de significación del 5%. Es decir, los *t-ratios*, en valor absoluto son superiores a 1,96 para cada parámetro estimado por tanto ambos coeficientes son estadísticamente significativos y deben permanecer en el modelo (ver Tabla 2). Cuanto a los modelos M_2 y M_3 , se comprueba que los coeficientes asociados a la componente AR(1) no permiten rechazar la hipótesis de nulidad del parámetro teórico; es decir, los valores del estadístico *t*, 1,15901 y 1,77011, para los modelos M_2 y M_3 , respectivamente, son insuficientes para rechazar su posible nulidad teórica, concluyéndose que estos coeficientes no son estadísticamente significativos, por lo que, y teniendo presente el principio de parsimonia, tales parámetros se deberían excluir de los modelos.

En cuanto a la posibilidad de que los modelos sean invertibles, analizando los tres modelos, las componentes (no estacional y estacional), verifican las condiciones de inversión, puesto que los estimadores de los parámetros de las componentes de medias móviles son, en módulo, inferiores a la unidad. Por naturaleza los procesos autorregresivos son invertibles. Se constata aún que dado que cualquier proceso de medias móviles es estacionario y estándose en presencia de valores estimados para el proceso autorregresivo, en los tres modelos, lejos de la frontera de la no estacionariedad, no es necesario diferenciar nuevamente la serie para concluir su estacionariedad.

El análisis de la matriz de las correlaciones de los coeficientes de un modelo puede ser indicativo sobre la necesidad de rechazar el mismo. En efecto, aunque los parámetros sean significativos, serán inadecuados si existe una fuerte correlación entre ellos, lo que afecta a la estabilidad del modelo, tornándolo inadecuado para la predicción (Murteira *et al.*, 1993).

Tabla 3: Matriz de correlaciones entre los parámetros, para cada uno de los modelos.

$M_1 = \text{ARIMA}(0,1,1) \times (0,1,1)_{12}$		$M_2 = \text{ARIMA}(1,1,1) \times (1,1,1)_{12}$				$M_3 = \text{ARIMA}(1,1,1) \times (0,1,1)_{12}$		
MA(1)	MA(12)	AR(1)	MA(1)	AR(12)	MA(12)	AR(1)	MA(1)	MA(12)
MA(1) $\begin{bmatrix} 1 & -0,05198 \end{bmatrix}$	MA(12) $\begin{bmatrix} & 1 \end{bmatrix}$	AR(1) $\begin{bmatrix} 1 & 0,76242 & 0,04569 & 0,14842 \end{bmatrix}$	MA(1) $\begin{bmatrix} & 1 & 0,01013 & 0,08039 \end{bmatrix}$	AR(12) $\begin{bmatrix} & & 1 & 0,69854 \end{bmatrix}$	MA(12) $\begin{bmatrix} & & & 1 \end{bmatrix}$	AR(1) $\begin{bmatrix} 1 & 0,76354 & 0,16772 \end{bmatrix}$	MA(1) $\begin{bmatrix} & 1 & 0,10451 \end{bmatrix}$	MA(12) $\begin{bmatrix} & & 1 \end{bmatrix}$

Observando la matriz de correlaciones (Tabla 3) entre los parámetros y, puesto que sólo correlaciones superiores, en valor absoluto, a 0,7 señalan parámetros fuertemente correlacionados se concluye que: existe poca correlación entre ellos para el modelo M_1 , conduciendo a la indicación de que es un modelo adecuado para describir el comportamiento

futuro de la serie temporal; lo mismo ya no se verifica para los modelos M_2 y M_3 , donde existe una fuerte correlación entre los estimadores AR(1) y MA(1), AR(12) y MA(12) y entre los estimadores AR(1) y MA(1), para los M_2 y M_3 , respectivamente. Estas correlaciones elevadas sugieren la mala calidad de los mismos, llevando a la conclusión de que los modelos no son los más apropiados para describir el comportamiento futuro de la serie temporal, por lo que ya se rechazan una vez que los estimadores de los parámetros de los modelos ponen en causa la estabilidad de los mismos, no se sometiendo a las siguientes análisis.

De notar que, no se ha realizado un análisis de redundancia en estos modelos, ya que la hipótesis de que los parámetros tomen el valor 1 esta muy alejada de los coeficientes estimados. Los modelos con redundancias o casi redundancias son inestables, dando origen a estimaciones de poca calidad y son contrarios al principio de parsimonia (Almargo, 1979).

Así, por lo presentado y analizado anteriormente, a continuación se sigue con el estudio del modelo M_1 , para observar su comportamiento en el resto de los criterios que aun restan por analizar.

Una vez evaluada la calidad estadística del modelo M_1 , se va analizar la calidad global del ajuste estudiando los respectivos residuos. Si el modelo explica correctamente la serie, los residuos estimados tienen que tener un comportamiento análogo al de un ruido blanco (*white noise*).

La Figura 9, representa el gráfico de los residuos del modelo M_1 . Se observa que se aprecian algunos atípicos en los años 1992, 1997, 2001 y 2002. También se observan fluctuaciones de un año para otro, en los meses de Marzo y Abril. Este comportamiento es debido, al menos en parte, al efecto de la festividad correspondiente a la Semana Santa.

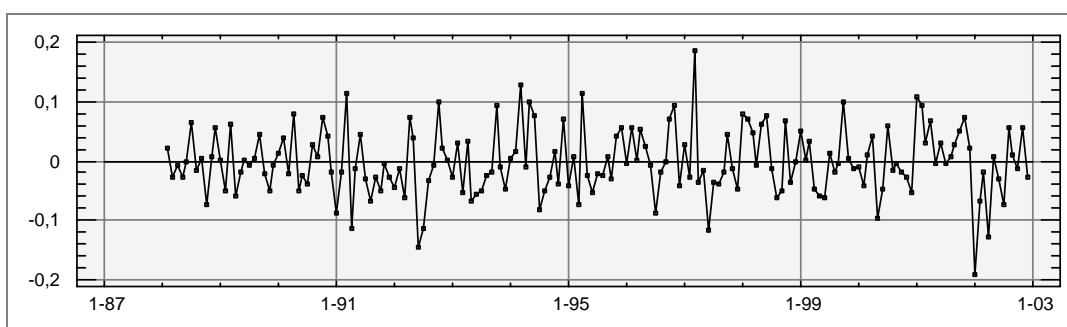


Figura 9: Gráfico de los residuos de M_1 .

Por lo que respecta a los residuos correspondientes a Julio y Agosto de 1992, su comportamiento puede estar justificado por la Guerra del Golfo y el correspondiente a los mismos meses de 1997, se pueden fundamentar tanto en la inestabilidad del mercado Ruso como en el conflicto de los Balcanes.

Entre Mayo y Noviembre de 2001, el comportamiento de los residuos correspondientes, puede estar justificado por el hecho de que la ciudad de Oporto haya sido la capital europea de la cultura, lo que produjo un aumento de los turistas, tanto nacionales como extranjeros, debido a los eventos culturales que tuvieron lugar con motivo de tal celebración.

En Diciembre de 2001, el casco histórico de la ciudad de Guimarães y el Alto Douro Vinhateiro, fueron clasificados por la UNESCO como Patrimonio Cultural Mundial, lo que sin duda despertó

la curiosidad de los turistas, esencialmente extranjeros, y llevo a desplazarse a mayor número de ellos a estos destinos turísticos. Este aumento de la demanda turística se vio reflejado en el año siguiente. Algunas exposiciones internacionales y cumbres internacionales (por ejemplo, la Organización para la Seguridad y Cooperación en Europa), realizadas en la ciudad de Oporto en el año 2002, fueron palco de polos de atracción turística y promoción turística conllevando a que los participantes, además de permanecer por más tiempo en el lugar que visitaban, se desplazaran a otras regiones limítrofes.

A partir del año 2000, una apuesta en una mayor promoción de la Región Norte y su imagen como destino de calidad, diferenciada y competitiva, frente a otras regiones y/o países, se ha visto reflejada en el comportamiento del turista nacional y extranjero, conduciéndolo a permanecer más en el lugar que visita. Esta promoción y divulgación por parte de las instituciones competentes se fundamenta esencialmente en la tradición, la artesanía, la cultura, la historia, la gastronomía, entre otras. También ha tenido su efecto el incremento en la divulgación de algunos destinos turísticos de extrema importancia para la región, como por ejemplo el Parque Natural del Douro Internacional y el Alto Douro Vinhateiro, la existencia de determinadas iniciativas culturales, recreativas y fiestas populares, en la promoción de ferias comerciales temáticas, que ha su vez han contribuido en la divulgación de algunos productos regionales, etc.

Una vez comprobada la adecuación de los residuos del modelo M_1 , se ha llevado a cabo el análisis global de los residuos por medio del estadístico de *Box-Pierce* para diferentes valores de los retardos. Observando los resultados de Tabla 4, comprobamos que los residuos obtenidos para el modelo M_1 estimado permiten aceptar la hipótesis de ruido blanco, pues el valor de los *p-value* asociados al contraste de *Box-Pierce* son elevados y bastante alejados del cero.

Tabla 4: Valores del Estadístico de *Box-Pierce*.

MODELOS	Q(12)	Q(24)	Q(36)
	[P-VALUE]	[P-VALUE]	[P-VALUE]
$M_1 = \text{ARIMA}(0,1,1) \times (0,1,1)_{12}$	7,74667 [0,653566]	13,5432 [0,916898]	27,1464 [0,791668]

Para terminar de construir el análisis, se ha recurrido a los criterios que miden la cantidad de información, concretamente a los criterios AIC^5 (-509,673) y BIC^6 (-503,298). Con efecto, los valores para los dos criterios son aproximados, concluyéndose que los mismos poseen comportamiento análogo.

En resumen, teniendo en cuenta los diferentes aspectos y criterios analizados para la evaluación de los distintos modelos, fue el M_1 , como ya se constató, el que presenta las mejores características teniendo en cuenta que:

⁵ $AIC(q) = n \cdot \ln \hat{\sigma}_a^2 + 2q$, Akaike Information Criterion.

⁶ $BIC(q) = n \cdot \ln \hat{\sigma}_a^2 + q(1 + \ln n) + q \cdot \ln \left[\frac{1}{q} \left(\frac{\hat{\sigma}_y^2}{\hat{\sigma}_a^2} - 1 \right) \right]$, Bayesian Information Criterion.

- a) los estimadores son significativamente diferentes de cero;
- b) las correlaciones entre los estimadores presentan valores razonables, pudiendo considerarse el modelo como estable, no así para los otros dos;
- c) los valores del estadístico de *Box-Pierce* permiten aceptar la hipótesis de que los residuos se comportan como un ruido blanco;
- d) finalmente, el modelo estimado y seleccionado podemos representarlo como:

$$(1-B)(1-B^{12})LSO_t = (1-0,628237B)(1-0,769101B^{12})e_t \quad [16]$$

$[t_1=10,5704] \quad [t_2=18,1643]$

es decir, los valores de la variable analizada, con sus ponderaciones indicadas, están influenciados por sus desfases en los periodos 1, 12 y 13 , así como de los residuos o errores en los mismos desfases y en el período actual. Se puede aún escribir de la siguiente forma:

$$\widehat{LSO}_t = LSO(-12) + LSO(-1) - LSO(-13) - 0,769101e(-12) - 0,628237e(-1) + 0,48317e(-13) + e_t \quad [17]$$

Sin embargo, la prueba concluyente del modelo elegido para captar el comportamiento de una serie es la predicción, en la cual se determina la efectividad del estudio realizado y se efectuará en la sección siguiente.

4. PREDICCIÓN DE LA DEMANDA TURÍSTICA EN LA REGIÓN NORTE DE PORTUGAL

Una vez seleccionado el modelo, $ARIMA(0,1,1) \times (0,1,1)_{12}$, que cumple satisfactoriamente los criterios estadísticos de evaluación establecidos y que mejor se adhiere a la evolución de las Pernoctaciones mensuales en el Norte de Portugal registradas en el período de Enero de 1987 a Diciembre de 2002, el paso siguiente es utilizarlo en realizar predicciones.

Las predicciones del modelo ARIMA, en función de los datos, vienen dadas por medio de la función de predicción:

$$Y_t^*(m) = E\{Y_{t+m} / Y_t, Y_{t-1}, Y_{t-2}, \dots\}. \quad [18]$$

El cálculo de la predicción $Y_t^*(m)$ puede hacerse de forma recursiva utilizando el modelo ARIMA estimado, de forma que si el modelo se escribe (Murteira *et al.*, 1993):

$$d_t = \phi_1 d_{t-1} + \dots + \phi_p d_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, \quad [19]$$

donde d_t es la diferencia de orden d de Y_t (supuesto Y_t no estacionaria y convertible en estacionaria mediante un proceso de diferenciaciones). Para calcular la predicción $Y_t^*(m)$, se comienza calculando la estimación de $d_t(1)$ como la esperanza condicionada de $d(t+1)$. Posteriormente se calcula la estimación de $d_t(2)$, también con la esperanza condicionada, y así sucesivamente hasta calcular la estimación de $d_t(m)$. Una vez que se tienen las predicciones de la serie d_t , se puede obtener una predicción de Y_t , teniendo en cuenta que

$$Y(t+m) - Y(t) = d(t+m). \quad [20]$$

Con ella se han obtenido los resultados de las predicciones para el año 2003. En la Tabla 5, se presentan las predicciones y los respectivos límites inferior y superior del intervalo de confianza al 95% construido para el modelo ARIMA. También se presenta, el error relativo absoluto de las predicciones, en porcentaje.

Tabla 5: Predicciones de las Pernoctaciones y el error relativo absoluto de las predicciones en %, para el período de 2003:01 a 2003:12.

MESES	LÍMITE INFERIOR	ARIMA	LÍMITE SUPERIOR	MAPE
Enero	150.315	175.679	201.673	13,0%
Febrero	156.283	184.154	213.667	3,6%
Marzo	215.031	248.721	277.865	16,2%
Abril	235.560	277.017	303.408	7,2%
Mayo	262.452	308.819	334.968	5,2%
Junio	266.683	316.477	343.584	16,6%
Julio	288.801	344.530	369.850	8,1%
Agosto	387.299	450.026	472.295	3,9%
Septiembre	313.724	372.961	402.490	8,6%
Octubre	247.638	301.144	340.022	7,0%
Noviembre	173.147	221.189	269.011	0,8%
Diciembre	137.817	188.072	237.040	5,4%
MAPE	-	-	-	7,9%

Nota: a) Error Porcentual Absoluto Medio $MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$.

Goh y Law (2002), citando a Lewis, señalan en relación a las precisión de las predicciones: cuando los errores relativos absolutos de las predicciones sean inferiores a un 10%, se considera una ‘precisión elevada’; entre 10 y 20%, se considera ‘buena’; entre 20 y 50% se considera ‘razonable’; y superiores a 50%, ‘imprecisas/poco fiables’.

De acuerdo con esta clasificación, existen algunos errores relativos absolutos superiores a 10% (los meses de Enero, Marzo y Junio); sin embargo, puede decirse que las predicciones pueden considerarse como fiables, dado que se registra una aplastadora mayoría de los errores relativos absolutos inferiores a 10%.

5. CONCLUSIÓN

El objetivo principal de este artículo, ha sido desarrollar un modelo lineal que permitiera efectuar predicciones de la demanda turística, para la Región Norte de Portugal. También se intentó, con la metodología adoptada, buscar una relación entre una modelización teórica sólida y el estudio práctico de la serie “*Pernoctaciones Mensuales de la Región Norte de Portugal*”.

Algunas de las desviaciones más notables están concentradas en algunos meses. Esto es debido a ciertos factores, que ya fueron comentados anteriormente, como el carácter móvil de la Semana Santa. Algunos acontecimientos en el Norte de España, como por ejemplo el

Xacobeo en Galicia, o la candidatura de Portugal al EUROPEO 2004 de fútbol, tienen sus repercusiones, tanto en los turistas portugueses como en extranjeros, que aprovechan parte de su tiempo para dedicarlo a conocer el norte de Portugal. Existen también otros factores como, por ejemplo, la inestabilidad del mercado del Oriente Medio y de la Europa del Este, que ha llevado a que algunos turistas nacionales permanecieran en Portugal y atrayendo a turistas extranjeros a disfrutar de la tradición, patrimonio cultural y arquitectónico, historia, gastronomía, etc, de la Región Norte de Portugal.

Así, lo que se ha pretendido en esta investigación, más que la obtención de un modelo identificado e invariable, ha sido definir un método que pudiera tener cierta flexibilidad una vez construido, de forma que sea capaz de incorporar fácilmente nueva información disponible y hacer nuevas predicciones futuras.

6. BIBLIOGRAFÍA

- ALMARGO**, José; (1979); "Aplicaciones del enfoque Box-Jenkins a series del Turismo Español"; *Cuadernos Económicos de Información Comercial Española*; N.º11/12, pp.53/107.
- CHU**, Fong-Lin; (1998); "Forecasting Tourist Arrivals: nonlinear sine wave or ARIMA?"; *Journal of Travel Research*; Vol. 36; pp.79/84.
- GOH**, Carey and **LAW**, Rob; (2002); "Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention"; *Tourism Management*; N.º23, pp.499/510.
- GREENE**, William H.; (2000); "*Econometrics Analysis*"; Prentice-Hall, Inc; Fourth Edition; USA.
- JOHNSTON**, Jack e **DINARDO**, John; (2000); "*Métodos Económicos*"; Editora McGraw-Hill de Portugal, Lda.
- KOHZADI**, Nowrouz; **BOYD**, Milton; **KERMANSHAHI**, Bahman and **KAASTRA**, Ieabeling; (1996); "A comparison of artificial neural network and time series models for forecasting commodity prices"; *Neurocomputing*; N.º10, pp.169/181.
- MAKRIDAKIS**, Spyros and **HIBON**, Michèle; (1997); "ARMA Models and the Box-Jenkins Methodology"; *Journal of Forecasting*; Vol. 16; pp.147/163.
- MURTEIRA**, Bento J.F.; **MÜLLER**, Daniel A. e **TURKMAN**, K. Feridun; (1993); "*Análise de sucessões cronológicas*"; McGraw-Hill; Lisboa.
- OTERO**, José Mª; (1993); "*Econometría - series temporales y predicción*"; Editorial AC; Madrid.
- PARRA**, S. B. y **DOMINGO**, J. U.; (1987); "Análisis de series temporales de turismo de la Comunidad Valenciana"; *Estadística Española*; N.º114, pp.111/132.
- PULIDO**, Antonio; (1989); "*Predicción Económica y Empresarial*"; Ediciones Pirámide; Madrid.
- WITT**, Stephen F. and **WITT**, Christine A.; (1995); "Forecasting tourism demand: a review of empirical research"; *International Journal of Forecasting*; N.º11, pp.447/475.
- WOLD**, Herman; (1938); "*A study in the analysis of Stationary Time Series*"; Almqvist and Wiksell; Upsala; 2d. Ed. 1954.

ANEXO A

Tabla A.1: Valores de la Serie Original, para el período entre 1987:01 y 1995:12.

MESES \ AÑOS	AÑOS								
	1987	1988	1989	1990	1991	1992	1993	1994	1995
ENERO	102.447	118.011	122.217	126.671	126.826	124.194	121.469	118.606	122.480
FEBRERO	102.123	117.547	116.837	129.802	131.653	127.474	129.284	122.988	130.393
MARZO	125.401	142.687	160.658	158.701	188.999	157.536	154.734	175.261	156.645
ABRIL	150.042	167.118	169.326	197.757	182.290	196.087	189.142	185.525	209.263
MAYO	180.430	189.823	199.158	207.876	219.187	223.918	198.402	232.075	218.666
JUNIO	197.113	207.729	218.595	227.159	251.295	207.907	207.216	248.237	222.720
JULIO	229.293	254.523	252.634	257.633	273.927	231.801	231.453	246.274	247.589
AGOSTO	304.847	315.113	329.014	351.500	341.490	312.026	304.576	322.366	320.750
SEPTIEMBRE	238.542	258.287	278.074	284.867	283.378	259.023	249.583	266.094	269.433
OCTUBRE	173.503	174.359	189.664	216.286	197.241	205.400	202.792	206.256	196.466
NOVIEMBRE	130.187	137.933	138.683	162.062	152.554	149.289	141.976	144.803	152.340
DICIEMBRE	114.229	128.774	127.730	139.683	132.802	130.963	120.748	139.706	140.643
TOTAL	2.048.157	2.211.904	2.302.590	2.459.997	2.481.642	2.325.618	2.251.375	2.408.191	2.387.388

Fuente: Elaboración Propia, basada en los datos del INE.

Tabla A.1: Valores de la Serie Original, para el período entre 1996:01 y 2003:12.

MESES \ AÑOS	AÑOS							
	1996	1997	1998	1999	2000	2001	2002	2003
ENERO	126.910	140.430	148.218	163.696	162.389	176.690	165.653	155.527
FEBRERO	139.403	141.183	157.415	165.988	162.637	186.586	181.005	177.818
MARZO	172.393	219.465	209.929	228.149	226.010	245.261	249.214	214.106
ABRIL	213.973	224.382	232.767	242.744	262.865	291.395	253.274	258.519
MAYO	239.142	253.833	280.326	269.854	264.497	306.743	302.028	293.531
JUNIO	245.264	238.334	296.612	270.126	273.881	325.568	301.465	271.454
JULIO	248.398	266.993	303.866	306.031	324.962	351.955	314.560	318.706
AGOSTO	336.086	345.672	377.645	385.868	397.405	452.581	444.991	433.211
SEPTIEMBRE	280.769	288.409	309.700	321.248	331.155	383.793	361.181	343.534
OCTUBRE	225.734	232.052	263.522	280.597	263.217	319.417	287.383	281.472
NOVIEMBRE	175.438	166.835	180.796	193.062	186.445	238.925	221.910	219.463
DICIEMBRE	143.163	141.349	161.273	166.990	157.210	202.351	179.766	178.439
TOTAL	2.546.673	2.658.937	2.922.069	2.994.353	3.012.673	3.481.265	3.262.430	3.145.780

Fuente: Elaboración Propia, basada en los datos del INE.